Natural image statistics and visual processing

van der Schaaf, Arjen

Chapter 4.1*

Independent Component Filters of Natural Images Compared with Simple Cells in Primary Visual Cortex

SUMMARY

Properties of the receptive fields of simple cells in macaque cortex were compared with properties of independent component filters generated by independent component analysis (ICA) on a large set of natural images. Histograms of spatial frequency bandwidth, orientation tuning bandwidth, length and aspect ratio of the receptive fields match well. This indicates that simple cells are well tuned to the expected statistics of natural stimuli. There is no match, however, in calculated and measured distributions for the peak of the spatial frequency response: the filters produced by ICA do not vary their spatial scale as much as simple cells do, but are fixed to scales close to the finest ones allowed by the sampling lattice. Possible ways to resolve this discrepancy are discussed.

INTRODUCTION

In this Chapter we investigate to what extent the statistical properties of natural images can be used to understand the variation of receptive field properties of simple cells in the mammalian primary visual cortex. The receptive fields of simple cells have been studied extensively (e.g., Hubel & Wiesel, 1968; DeValois *et al*., 1982a; DeAngelis *et al*., 1993): they are localised in space and time, have band-pass characteristics in the spatial and temporal frequency domains, are oriented, and are often sensitive to the direction of motion of a stimulus. Here we will concentrate on the spatial properties of simple cells.

Several hypotheses as to the function of these cells have been proposed. As the cells preferentially respond to oriented edges or lines, they can be viewed as edge or line detectors. Their joint localisation in both the spatial domain and the spatial frequency domain has led to the suggestion that they mimic Gabor filters, minimising

* The work presented in this Chapter has been accepted for publication by the *Proc. R. Soc. Lond. B.* The ICA algorithm was implemented by van Hateren, the extraction of parameters from IC-filters was implemented by van der Schaaf.

uncertainty in both domains (Daugman, 1980; Marcelja, 1980). More recently, the match between the operations performed by simple cells and the wavelet transform has attracted attention (e.g., Field, 1993). The approaches based on Gabor filters and wavelets basically consider processing by the visual cortex as a general image processing strategy, relatively independent of detailed assumptions about image statistics. On the other hand, the edge and line detector hypothesis is based on the intuitive notion that edges and lines are both abundant and important in images. This theme of relating simple cell properties with the statistics of natural images was explored extensively by Field (1987, 1994). He proposed that the cells are optimized specifically for coding natural images. He argued that one possibility for such a code, sparse coding, simplifies further processing in the visual system, because it produces a representation of the stimulus that helps detection of coincidences (Barlow, 1972, 1994). Indeed, Olshausen and Field (1996) showed that imposing sparseness on the output of receptive fields being trained on natural images produced receptive fields similar to those of simple cells (see also Harpur, 1997).

This result was very recently put into the context of independent component analysis (ICA) by Bell and Sejnowski (1997a, 1997b) and Hurri *et al.* (1996). For (linear) ICA one considers an ensemble of signals, each produced by an unknown linear superposition of unknown independent (elementary) signals. By presenting the ICA algorithm with a large number of examples of such signals, it is able to reconstruct the elementary signals, at least if the elementary signals have non-gaussian probability densities (i.e., the distribution of the strengths with which each elementary signal is present in a set of images is not a Gaussian). ICA on natural images again produces receptive fields like those of simple cells (Bell & Sejnowski, 1997a,b; Hurri *et al.*, 1996; Hurri, 1997). Although the components produced by ICA on natural images are not completely independent, they are as independent as possible by a linear transformation.

It should be noted here that the independent component model of the primary visual cortex should not be regarded as a full model of simple cells in primary cortex. As it is a linear, non-adaptive model, many aspects of simple cells are ignored, such as contrast adaptation (e.g., Sclar *et al.*, 1989), contrast normalisation (Heeger, 1992), nonlinearities involved in orientation tuning (Volgushev *et al.*, 1996), adaptation to various stimulus statistics (Zipser *et al.*, 1996), and so on. Nevertheless, the model has a clear information theoretic interpretation, and is strictly based on stimulus statistics. One may hope that nonlinearities and adaptation can be added in a process of stepwise refinement, with the linear model as a solid basis. This appears to be a successful strategy in earlier stages of vision (vertebrate retina and LGN, fly visual system), where linear theories are quite successful (e.g., Srinivasan *et al.*, 1982; Atick, 1992; van Hateren, 1992a,b; Linsker, 1993; Dong & Atick, 1995b; Dan *et al.*, 1996), whereas it is known that the early visual system contains (e.g., Laughlin, 1981; van Hateren 1997) and needs (e.g., Ruderman & Bialek, 1994) stimulus related nonlinearities that modify the linear default.

If decomposing an image into independent components is indeed one of the main functions of simple cells, it is expected that the distribution of their properties, such as spatial frequency bandwidth and orientation tuning bandwidth, is determined by

the statistics of the visual environment. Olshausen and Field (1996) report spatial frequency bandwidths and aspect ratios (ratio of length and width of the receptive field) close to those measured in simple cells. Here we extend this result by performing ICA on a large set of calibrated images, and comparing a series of properties of the resulting receptive fields with those of receptive fields measured in simple cells. We find that there is a good correspondence between the distributions for spatial frequency bandwidth, orientation tuning bandwidth, aspect ratio, and receptive field length. For the peak of the spatial frequency sensitivity, however, the results deviate strongly: whereas ICA yields peaks close to the maximum spatial frequency allowed by the sampling lattice, measurements in simple cells show a much broader distribution. The implications and possible resolution of this discrepancy are discussed.

METHODS

Images

The image set consisted of 4212 images obtained with a Kodak DCS420 digital camera (with a 28 mm camera lens). For the intensity this camera uses 12-bit sampling internally, which is then reduced to and stored as 8-bit data via a nonlinear scale table. As this table is recorded for each image, it can be used afterwards to expand the 8-bit data to a linear scale. Although the latter scale is, strictly speaking, not genuinely 12 bit deep, it is effectively close to it: details in the shadows are retained, whereas high peaks of intensity (see, e.g., van Hateren, 1997) are more roughly quantised, but without the clipping that an 8-bit system would necessitate. The linearity of the camera was checked with a set of calibrated neutral density filters, and found to be satisfactory. The (slight) blur in the images caused by the optical system of the camera was measured and corrected through the following procedure (performed separately for each diaphragm of the camera lens used). Spatial point spread functions of the camera were measured for a small point source presented at a large number of random positions. The resulting images (sampled point spread functions) were Fourier transformed, resulting in slightly different amplitude spectra depending on the exact position of the image of the point source on the CCD sampling grid. Of these spectra, the group with the shallowest high frequency fall-off was selected for inverse filtering the images of natural scenes. Only the images of point sources corresponding to those spectra would thus be reconstructed as a sharp point source. Point sources at other positions with steeper spectra would remain slightly blurred in a reconstruction. As a result of this procedure, the deblurring does not cause spurious structures (such as fringes at the edges) in the images of natural scenes. After inverse filtering, the images had a resolution of 1536x1024. This was subsequently reduced to 768x512 by block averaging, before extracting image patches for the ICA. This procedure reduces the risk of any remaining calibration insufficiencies influencing the ICA. Noise in these images is negligible. The final images had an angular resolution of approximately 2 minutes of arc per pixel. The images were taken in various environments (wood, open landscapes, urban areas). In a particular environment a typical series of 100-200 consecutive images was taken. Therefore, when going through the series of 4212

images, the images statistics may change regularly. Some of this variability was incorporated in the results of the ICA by using samples from consecutive images rather than random ones taken from the entire set for a particular ICA run (see below).

A second set of images that was used (12x12 video in Fig. 4.1.3) consisted of frames grabbed from television broadcasts. This set was uncalibrated, but more diverse than the calibrated set. Images were taken from programs on a wide range of subjects, including wild life (ranging from arctic to tropical), sports, and movies.

Independent Component Analysis

Algorithms performing (linear) ICA decompose each signal of an ensemble into components (also called 'basis vectors') that are as independent as possible by a linear transformation of the signals. The amplitude of a particular component is extracted by a corresponding weight vector (also called a 'filter', see Bell & Sejnowski, 1997). Of the various algorithms for performing ICA, the recent algorithm presented by Hyvärinen and Oja (1997a,b) and Hyvärinen (1997) was used. This algorithm implements ICA by finding filters that produce extrema of the kurtosis. Note that either maxima or minima in the kurtosis can be found (making it more general than maximising sparseness), although we found in practice only maxima for our data set. The idea of the method is that the independent components must have an extremum of the kurtosis, as any impure component (i.e., a linear superposition of two or more pure independent components) would result in a change of the kurtosis towards zero (in the limit a kurtosis of zero, as the central limit theorem states that a linear summation of a large number of independent variables, each with finite variance, will produce a Gaussian distribution, which has zero kurtosis). The Hyvärinen algorithm (the fixed-point algorithm, with a serial deflation scheme as in Hyvärinen & Oja, 1997b, using function g_2 as in Hyvärinen, 1997) was implemented as a parallel program on a Cray J932 at the Centre for High-Performance Computing of the University of Groningen. A typical run was performed on 120000 different samples of 18x18 image patches taken evenly spaced from 100-120 consecutive images of the image data set; such a run took about half an hour. Histograms were compiled from 33-160 of such runs, based on different subsets of the image set.

Before processing, the logarithm of the intensities was taken, firstly, because this incorporates the contrast invariance of natural scenes, secondly, because it leads to better behaved first-order statistics of natural images (e.g., Ruderman, 1994), and, thirdly, because it is similar to the operations performed by the first stages of visual systems (e.g., van Hateren, 1997). Note, however, that for the low contrast stimuli as typically used for measuring simple cell properties (e.g., the histograms in Fig. 4.1.4) taking the logarithm hardly makes any difference. The results of the ICA algorithm were obtained in PCA whitened space (see below), and subsequently transformed back to the original image space (with intensities represented on a logarithmic scale). All results presented below are in this space, and the filters (e.g., Fig. 4.1.1B) are thus meant to work on the (logarithmic) original images.

Early calculations were performed without dimensionality reduction in the PCA (the principal component analysis which is part of the data whitening used before applying the Hyvärinen algorithm). This leads to about 70% oriented IC filters, and 30% non-oriented. The latter only extend over a few pixels, and have about equal power in all four corners of the power spectrum. We consider these filters as an aliasing artefact, and they were not included in the analysis (by selecting only the filters with oriented power spectra). In later calculations, the dimensionality of the data was reduced by about 25% (i.e., from 18×18=324 degrees of freedom to 240 degrees of freedom, by selecting the most significant principal components), which leads only to oriented filters; these were all included in the analysis. Both of the above procedures led to very similar histograms for all parameters investigated. Also reducing the dimensionality further to 160, 80, or 40 degrees of freedom led to similar results, only differing in the spatial scale, but not in the relative distribution of the parameters.

RESULTS

Figure 4.1.1 shows an example of independent components (Fig. 4.1.1A), i.e., the basis vectors, and the corresponding filters required to extract the strength of each component from an image (Fig. 4.1.1B), i.e., the weight vectors (or transform coefficients). Both basis vectors and filters are shown here (and below) as they appear in image space (after the logarithm of the intensity was taken). Whereas the independent components (ICs) can be considered as the constituting elements of images (see Bell & Sejnowski, 1997a,b; Olshausen & Field, 1996), the IC filters needed to extract their strengths are analogous to the receptive fields of neurons analysing a scene. Thus the receptive fields of cortical neurons should be compared to the IC filters, and not to the ICs. As discussed by Bell and Sejnowski (1997a,b), the ICs resemble short edges, and the IC filters are similar to simple cell receptive fields, i.e., resembling Gabor filters or wavelets. The filters are usually low-pass in one direction, and band-pass in the orthogonal direction. Figure 4.1.1C shows the amplitude spectra of the filters, with zero spatial frequency in the centre of each patch. From sets of filters as in Fig. 4.1.1B and 4.1.1C it is possible to extract several descriptive parameters. Following work analysing cortical receptive fields (in particular DeValois *et al*., 1982a,b; Parker & Hawken, 1988), we will investigate here the following properties:

- spatial frequency bandwidth; this is defined here as the full width at half maximum (FWHM) of each filter along the orientation of the peak in the amplitude spectrum; it is expressed in octaves (factors of two in frequency).
- orientation tuning bandwidth; defined as the FWHM along a circle (with its origin at zero spatial frequency) through the peak in the amplitude spectrum.
- peak spatial frequency and peak orientation: spatial frequency and orientation of the peak in the amplitude spectrum.

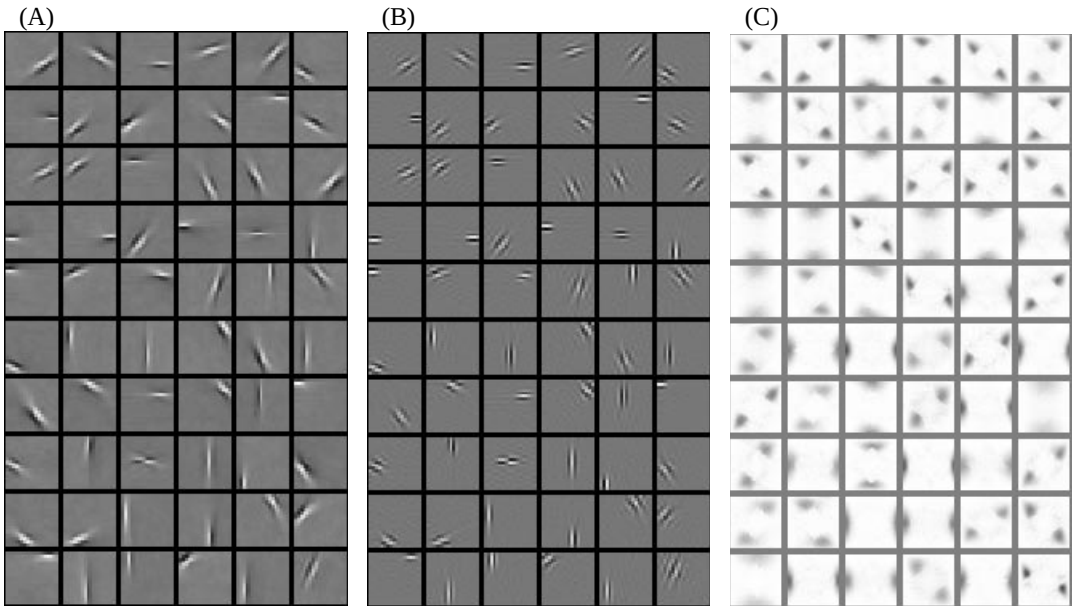

FIGURE 4.1.1: Independent component analysis on natural images (18×18 patches from 103 consecutive images, dimension reduced via PCA to 240 of 324). (A) IC basis vectors, and (B) corresponding IC filters (filtering an image with an IC filter yields the strength of the corresponding basis vector in the image). Signs of basis vectors and filters are arbitrary. (C) Amplitude spectra of the filters of (B), with darker grey values coding larger amplitudes. Zero spatial frequency is at the centre of each patch.

- length and aspect ratio of the receptive fields; first the (2D) Hilbert transform of the filter was calculated (the quadrature phase filter), which then yields, together with the original filter, the filter's envelope by subsequently squaring both filters, adding the results, and taking the square root (see Field & Tolhurst, 1986; DeAngelis *et al.*, 1993). The length is then defined as the FWHM of the envelope along the orientation into which the filter is low-pass, and the width as the FWHM along the orientation into which the filter is band-pass. The aspect ratio is defined here as the ratio of length and width.

We found that the distribution of each parameter depends on several factors. First, it depends on the particular set of images used for the ICA. Figure 4.1.2A shows several examples of the distribution of orientation tuning bandwidths for different subsets of images. Each of the curves resulted from ICA on a different set of 103 consecutive images taken from the entire set. All 40 sets thus analysed had histograms similar to or scattering between the ones shown in the figure. To further illustrate the variability of both the IC basis vectors and the IC filters, Fig. 4.1.2B-E shows samples of these (representative of the complete basis set) for the two subsets corresponding to the thick lines in Fig. 4.1.2A. The left thick line corresponds to basis vectors in Fig. 4.1.2B and filters in 4.1.2C, the right thick line to 4.1.2D and 4.1.2E. The first image set consists mostly of images of a waterside landscape with grass and reed, the second image set consists mostly of images inside a wood, with trees, leaves, and foliage. Observed superficially, the two sets do not appear very much different, but as Fig. 4.1.2B-E shows they lead to rather different sets of ICs. This suggests that the number of independent components needed for a large

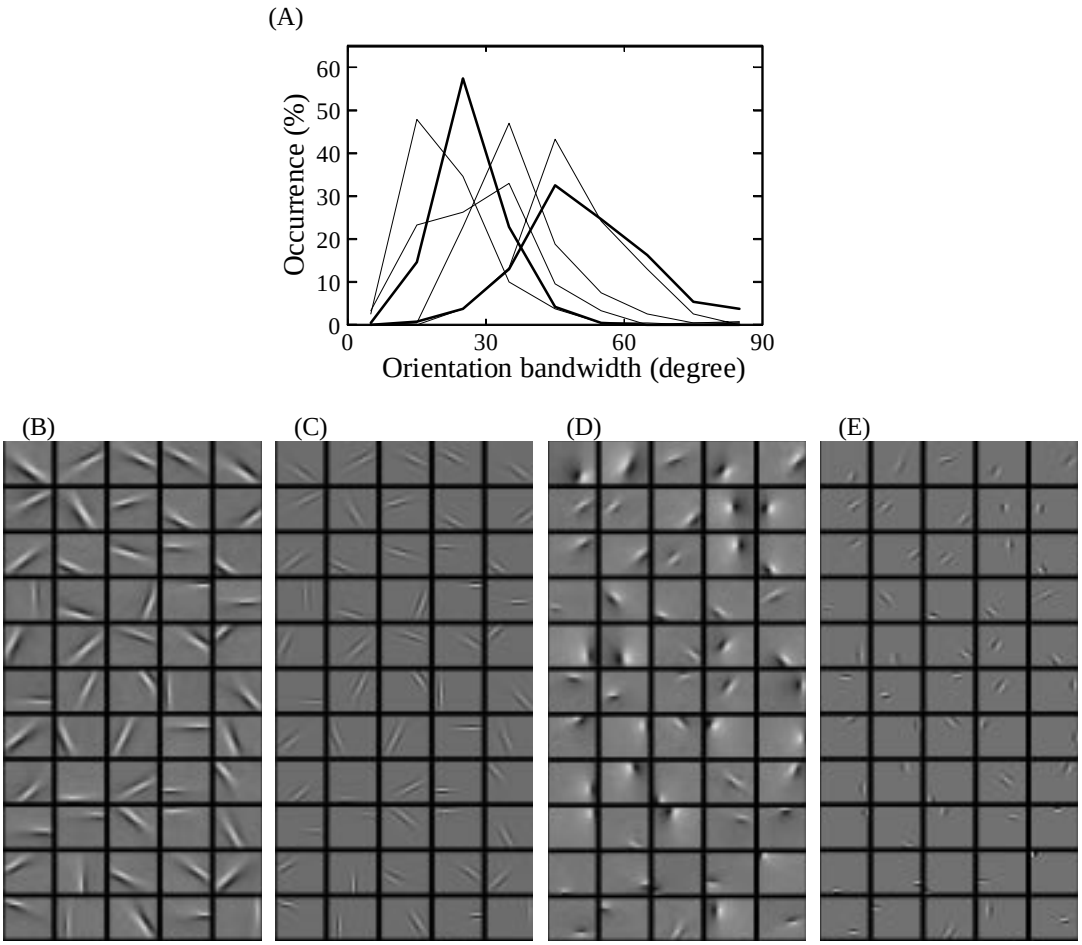

FIGURE 4.1.2: Variation of ICs for different ensembles of images. (A) Distribution of orientation tuning bandwidth for 6 different ensembles. (B) Examples of basis vectors for the left thick curve in (A), and (C) corresponding IC filters. (D) Examples of basis vectors for the right thick curve in (A), and (E) corresponding IC filters. Signs of basis vectors and filters are arbitrary. See text for further details.

ensemble of images is larger than the number of degrees of freedom of single image patches (see also the Discussion).

Variability of the results of ICA is further investigated in Fig. 4.1.3 for the spatial frequency bandwidths of the IC filters. Here we first varied the way of extracting image patches from the database: either patches taken from consecutive image sets (18×18 consecutive as above), or patches drawn randomly from the entire database (18×18 random). The latter procedure still yielded a wide range of bandwidths, but somewhat more restricted than for the consecutive image sets. Decreasing the size of the consecutive image sets from 103 images per set to 26 per set (18×18 short) did not noticeably broaden the distribution further. Second, we varied the type of preprocessing (PCA reduced, with the dimensionality of the patches reduced from 324 to 240), and the size of the patches considered (to 12×12). This produced similar histograms as the 18×18 consecutive results. Finally, an entire different image database was used (12×12 video images generated from video frames grabbed from television broadcasts). Again, the results are not dramatically different.

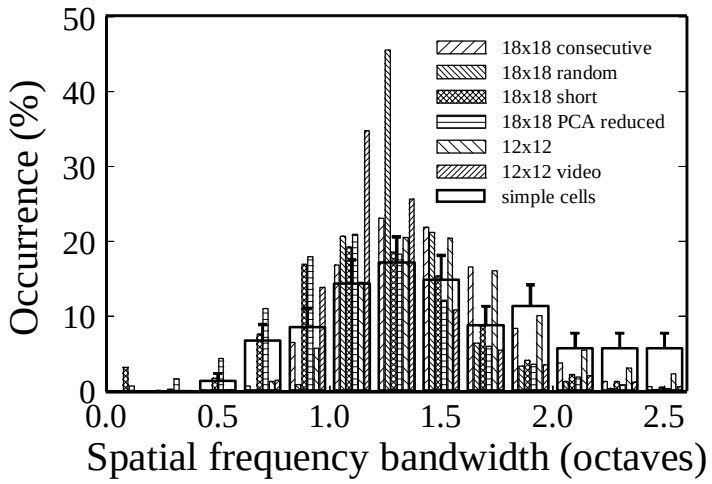

FIGURE 4.1.3: A comparison of different ways to perform the ICA with each other and with measurements in macaque simple cells. Data from DeValois *et al*., 1982b, their figure 5, foveal simple cells. See text for further details.

Figure 4.1.3 also shows measurements of spatial frequency bandwidths (FWHM) in simple cells in macaque cortex (histogram bars, data from DeValois *et al*., 1982b, cells recorded from the foveal area). Comparing these measurements with the ICA results, we see that the distributions roughly match, despite the variability in ICA results. This variability should probably not be considered as just a random estimation error, but rather as an inherent property associated with systematic changes in the statistics of different ensembles of natural images. Therefore, we consider the standard deviation of the ICA histograms (rather than the standard error of the mean) as a rough but reasonable measure for the reliability of the estimate. Below we will use this, together with the mean of the ICA histograms, for comparing the ICA results with various properties measured in simple cells.

Figure 4.1.4A shows this for the same data as in Fig. 4.1.3. The error for each bin of the histogram of simple cells was estimated by taking the square root of the number of cells in each bin. Figure 4.1.4A shows that the calculated and measured distributions are roughly similar; a chi-square test yields $p>0.26$, which means that the hypothesis that the calculated and measured histograms are identical can not be rejected. The mean and the standard deviation of the spatial frequency bandwidths shown here are similar to those reported by Olshausen and Field (1996), although a direct comparison is not fully appropriate because they analyse basis functions rather than filters.

Figure 4.1.4B shows that calculated and measured orientation tuning bandwidths are similarly distributed (data from both DeValois *et al*., 1982a and Parker & Hawken, 1988). Although the calculated curve appears to peak at a slightly smaller orientation tuning bandwidth (20°-30°) than the measurements in simple cells (30°-40°), the errors in the curves are such that they are not significantly different ($p>0.18$).

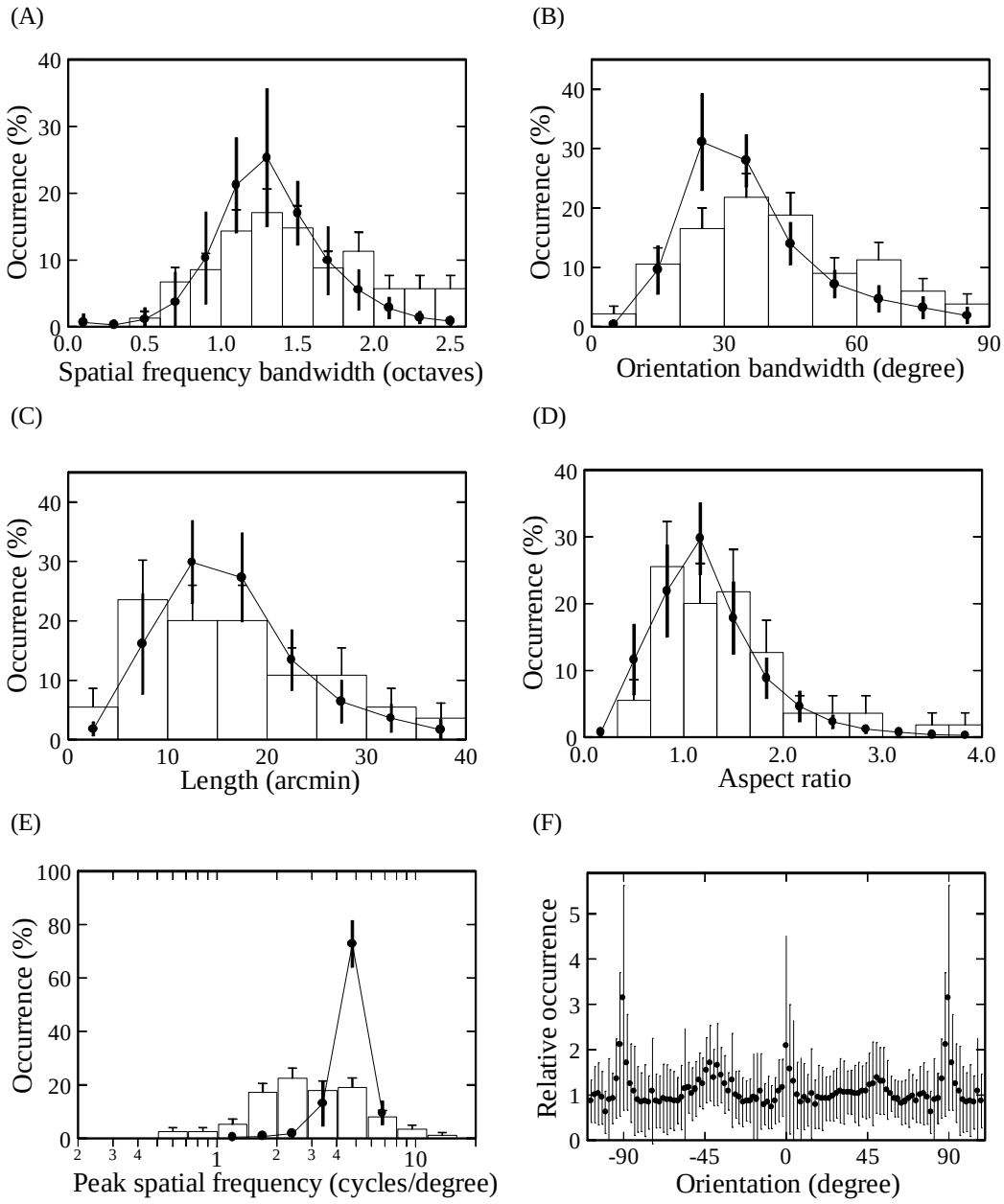

FIGURE 4.1.4: Comparison of IC filters (dots with error bars show mean and s.d. of the calculated histograms as in Fig. 4.1.3) to measurements of properties of simple cells in the foveal area of macaque primary visual cortex (histograms). Data from DeValois *et al.*, 1982a [(B), their fig. 3A; merged with data from Parker & Hawken, see below], DeValois *et al*., 1982b [(A), fig. 5; (E), fig. 6], and Parker & Hawken 1988 [(B), fig. 3a; (C), fig. 4a; (D), fig. 6a]. Scaling in (C) for calculated data: 1.2 pixel/bin. Filters sensitive to horizontal structures correspond to 0° in (F). See text for further details.

Figure 4.1.4C shows distributions of the length of the calculated filters and measured receptive fields ('height' as measured by Parker & Hawken, 1988). Here the abscissa gives the length (in minutes of arc of visual angle) for the histogram of simple cell measurements. The calculated curve was scaled along this axis such that it provided a good match with the measurements. Thus only the shape of the distributions should

be compared, not the position of the peaks. This procedure was necessary, because the calculations do not yield absolute visual angles that can be compared directly with those of the primate visual system: firstly, because the spatial resolution of the first data set used (taken with the digital camera) is not as fine as in the primate fovea (2′ between pixels for the image patches, versus approximately 30″ for the primate fovea); secondly, because, even if they had been identical, the true sample base as used by the primate cortex may be different from that of the cones; and thirdly, because the spatial calibration of the second data set (images grabbed from television broadcasts) is not known and variable because of diverse camera lenses. If images are approximately scale invariant over the relatively small range of scales involved here, the distribution of lengths should be scale invariant as well. Figure 4.1.4C shows that the shape of the calculated and measured distributions of receptive field lengths is similar ($p>0.68$).

The aspect ratio for the calculated filters is defined here as the ratio of length and width of the envelope of the filter (see above). From measurements by Parker and Hawken (1988) of length and width of the central lobe of the simple cell receptive field, the aspect ratio was estimated by assuming that the width of the envelope is about three times the width of the central lobe. Figure 4.1.4D shows the resulting distributions: again a close correspondence between ICA filters and receptive fields in simple cells ($p>0.47$). Similar aspect ratios were reported by Olshausen and Field (1996) for their basis functions.

Figure 4.1.4E shows the distribution of spatial frequencies giving the maximum response for ICA filters and for simple cells (data from DeValois *et al.*, 1982b). Contrary to the results presented above, the two distributions now deviate strongly ($p<<0.001$). Again there is no well-defined spatial scale for the calculated curve, which was arbitrarily positioned. As will be clear from the figure, any other positioning (e.g., aligning the peaks of the distributions) would also give a strong mismatch. The main reason for the mismatch is the tendency for the ICA to produce filters at a scale as close as possible to the sampling grid of the images (see the Discussion).

This property of ICA also causes an excess of filters aligned with the sampling lattice, i.e., horizontally and vertically. Thus a histogram of peak orientations yields a broad distribution with large and sharp peaks at horizontal and vertical orientations. These peaks can not be attributed directly to horizontal and vertical structures in the images, because they also occur when each image patch of the set of 120000 used for the ICA is rotated to a random orientation. In that case no inherent orientation can be present in the set of image patches, whereas there are still peaks in the results at horizontal and vertical orientations, be it somewhat smaller than before. In order to study possibly inherent orientational biases in the image set, despite the tendency of the ICA to align with the sampling lattice, we used the following procedure. We performed a series of ICA runs on 1) image patches all rotated over a fixed angle (30° for the results shown here, similar results were obtained with 45° and 22.5° rotations), and 2) image patches each rotated over a random angle. Set 1 is expected to yield peaks at horizontal and vertical orientations (the artefact), plus, e.g., a peak shifted by 30° from the horizontal if the image set has inherently more horizontal

structures. Set 2 is expected just to yield the artefacts at horizontal and vertical orientations. By taking the ratio of the histograms of orientations for set 1 and set 2, the excess at horizontal and vertical orientations in the ICA filters is expected to cancel, at least approximately. Any remaining anisotropy at other orientations must be caused by true anisotropy in the images.

The result of this operation shows peaks at 30° and -60°, i.e., not aligned with the square sampling lattice. Rotating all orientations by -30°, back to the original upright position of the images, produces Fig. 4.1.4F. Dots and error bars show the mean and standard deviation of 40 different image sets. As can be seen, the IC filters are relatively more often oriented horizontally (0°) and vertically (90°) than in other directions, but the effect is so small compared to the variability between subensembles of images that one could easily fit a horizontal line to the data. The width of the horizontal and vertical peaks is remarkably small; this is apparently caused during photography, by habitually aligning the camera frame with dominant horizontal and vertical structures in the image. Thus it depends on how precisely the human eye aligns with such structures during natural vision whether these peaks have any functional significance for human vision. The small peaks at about ±45° may also arise from a photographer's bias: one possible cause is that straight lines in the landscape (such as a road or path, or the fringe of a wood) are often framed diagonally, because this appears to produce pleasing, balanced pictures.

DISCUSSION

The independent component model is at present probably one of the most sophisticated ecologically inspired models for understanding the image representation in the array of simple cells in the human primary cortex. Not only does it produce quantitative predictions of receptive fields which compare reasonably well with those measured, it also gives a functional interpretation of the visual processing performed in terms of information theory and natural image statistics. In this Chapter we have put this model to the test by applying it on a large set of images, and comparing the distribution of predicted properties to those measured in simple cells. We showed that most of these distributions are similarly shaped. This applies to spatial frequency bandwidth, orientation tuning bandwidth, length of the receptive fields, and aspect ratio. This result implies two conclusions. Firstly, it strengthens the hypothesis that cortical simple cells strive to produce a representation of natural images with independent variables, each having a highly kurtotic amplitude distribution (i.e., with long tails, leading to sparse coding). Secondly, it suggests that the apparent randomness of simple cell properties may not be the sign of a sloppy design, nor of random variability in development, but may in fact be a deliberate attempt to match the requirements of processing natural images.

A notable exception to the correspondence between IC filters and simple cells is the distribution of the peak of the spatial frequency response. Whereas simple cells have receptive fields acting on different spatial scales (i.e., they show spatial scaling), the IC filters show much less variability (Fig. 4.1.4E). This discrepancy may be resolved in several ways. One possibility is that spatial scaling should be imposed as an extra

constraint (see, e.g., Koenderink, 1984; Li & Atick, 1994), as it can be argued that spatial scaling is a useful property for higher visual processing (e.g., object recognition relatively independent of viewing distance). Another possibility is that focussing on high spatial frequencies is a property of the particular ICA algorithm used (or even any algorithm maximising kurtosis, see Baddeley, 1996). However, ICA performed by Bell and Sejnowski (1997a,b), with a different algorithm, yielded similar results as reported here. The basis vectors calculated by Olshausen and Field (1996, 1997) show a somewhat higher occurrence of low spatial frequency ones, but still considerably less than measured in simple cells (cf. Fig. 4.1.4E). A further possibility is that extensions of the model to nonlinear ICA, possibly with overcomplete bases (i.e., more basis vectors than degrees of freedom), may resolve this discrepancy purely in the spatial domain. More likely, however, it will be necessary to include the time domain in the analysis. A preliminary analysis of the results of ICA performed on video sequences shows that spatiotemporal IC filters with peaks at low spatial frequencies are abundant, and are associated with higher speeds in the visual scenes.

Linear ICA yields the same number of basis vectors and filters as the number of degrees of freedom of the input. This is far less than the number of cells in the cortex per independent viewing direction as coded by the optic nerve. Part of this divergence may arise in ICA as well when including the time domain, leading to cells with different temporal frequency and velocity tuning. That there is also a need for divergence purely in the spatial domain is suggested by Fig. 4.12B-E: different ensembles of images need different sets of IC filters. Current work on overcomplete bases (Olshausen & Field, 1997; Lewicki & Olshausen, 1998) may lead to similar variability when performed on a sufficiently large and diverse image database.

Chapter 4.2

Edges, Local Contrast, and the Independence of Complex Cell Responses

INTRODUCTION

The second order statistics of natural images are enough to predict the response properties of retinal ganglion cells (e.g., Srinivasan *et al.*, 1982; Atick & Redlich, 1992). These statistics do not, however, capture all of the structure that is present in natural images. Most remarkably, second order statistics are blind for local orientation (Zetzsche *et al.*, 1993). This means that edges, which are typical features of natural images, are not described effectively by the second order statistics. The local structure of an edge forms an intrinsically one-dimensional structure that is redundant in the direction along the edge. Describing this redundancy is important to understand the response properties of cortical complex cells, which are known to respond preferably to edges of a specific orientation (Hubel & Wiesel, 1968).

We investigate the statistics of edges with the use of a trick: We transform the image in a way that explicitly extracts the edges of a specific orientation. This is done by analogy of cortical complex cells. This analysis will quantify the statistics of edges in natural images and, at the same time, show what edges imply to the coding properties of complex cells. An important property of image representation in the cortex is independence (see Chapter 1 and Chapter 4.1). We will, therefore, focus on correlation and mutual information between individual units.

METHODS

Cortical complex cells respond to images with the following properties: They are bandpass, orientation selective, and localised. A linear transform with these properties is the Gabor representation, which resembles the response properties of cortical simple cells (Daugman, 1980). A difference between simple cells and complex cells, however, is that complex cells respond always excitatory to edges. This is a desirable property for the present analysis of edges, since it identifies edges by a positive response, while simple cells give a modulated response. This property of

complex cells is modelled by taking the square modulus of the output of a complex analytic linear filter (Daugman, 1993). The impulse response of the analytic filter is given by:

$$h(\mathbf{x}) = e(\mathbf{x}) + i \cdot o(\mathbf{x}) \tag{4.2.1}$$

where the real part, $e(\mathbf{x})$, is an even function of the two-dimensional spatial coordinate ($\mathbf{x}$), and the imaginary part, $o(\mathbf{x})$, is an odd function. The response, $a(\mathbf{x})$, of a complex cell at spatial position $\mathbf{x}$ to an image, $g(\mathbf{x}')$, is then given as:

$$\begin{aligned} a(\mathbf{x}) &= \left|h * g(\mathbf{x})\right|^2 = \left|e * g(\mathbf{x})\right|^2 + \left|o * g(\mathbf{x})\right|^2 = \\ &= \left|\sum_{\mathbf{x}'} e(\mathbf{x}-\mathbf{x}') \cdot g(\mathbf{x}')\right|^2 + \left|\sum_{\mathbf{x}'} o(\mathbf{x}-\mathbf{x}') \cdot g(\mathbf{x}')\right|^2, \end{aligned} \tag{4.2.2}$$

where $*$ denotes convolution. The image value $g(\mathbf{x}')$ represents the logarithm of the intensity at spatial position $\mathbf{x}'$. Taking the logarithm of the intensity results in a representation of contrast, i.e., relative intensity difference, rather than absolute intensity difference.

The convolution $h*g(\mathbf{x})$ is more conveniently calculated in the Fourier domain as the product of the frequency response, $H(\mathbf{f})$, of the analytic filter and the Fourier transform, $G(\mathbf{f})$, of the image, $g(\mathbf{x})$. A problem with calculation in the Fourier domain is the image periodicity, which causes boundary artefacts. Therefore, we ignore the complex cell responses at the border of the image. An extra advantage of calculation in the Fourier domain is that the analytic filter, $h(\mathbf{x})$, of Eq. 4.2.1 is more conveniently described by its frequency response, $H(\mathbf{f})$. The bandpass selectivity and orientation tuning properties of cortical complex cells imply that the frequency response of the analytic filter is bandpass along the preferred orientation axis and lowpass in the perpendicular direction. Here these properties are modelled by the function

$$H(\mathbf{f}) = |\mathbf{f}| \cdot e^{\frac{-(f_1 - f_0)^2}{2\sigma_1^2} + \frac{-f_2^2}{2\sigma_2^2}}, \tag{4.2.3}$$

with $\mathbf{f}$ the two-dimensional frequency vector, and

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \cdot \mathbf{f} \,. \tag{4.2.4}$$

This analytic filter is the product of a whitening filter and a Gabor filter. The whitening filter, $|\mathbf{f}|$, flattens the $1/f^2$ power spectrum of natural images (see Chapter 2), and also removes the DC response. The Gabor filter is a bandpass filter that is tuned to an optimal frequency f_0 and orientation φ. The bandwidth in the two directions is controlled by the parameters σ_1 and σ_2. The filtered image has a two-dimensional spatial parameter $\mathbf{x}$, which gives each modelled complex cell six free parameters. Our results, however, do not critically depend on the bandwidth parameters σ_1 and σ_2 and we simplify the analysis by setting their values to typical values that are found for complex cells: $\sigma_1 = 0.54\, f_0$, and $\sigma_2 = 0.34\, f_0$. These values correspond to a frequency bandwidth of 1.5 octaves (De Valois, *et al.*, 1982b) and an

orientation bandwidth of 38° (De Valois, *et al.*, 1982a). These parameter settings fix the shape of the receptive fields $e(\mathbf{x})$ and $o(\mathbf{x})$. Their scale is controlled by the parameter f_0, which is set to 0.20 cycles/pixel, when not stated otherwise. With this parameter setting, the envelope of the receptive field, $|h(\mathbf{x})|$, has an effective size of 4.4 pixels in the preferred orientation and 3.4 pixels in the perpendicular orientation. These widths are calculated as two times the root mean square distance from the centre of the envelope profile along the axis. The resulting impulse response of the analytic filter (Eq. 4.2.1) is found by inverse Fourier transforming Eq. 4.2.3. An example of the real and imaginary receptive fields with typical parameters is shown in Fig. 4.2.1.

The set of complex cells that are tuned to the same bandwidth, scale, and orientation is called a channel. Examples of the responses, $a(\mathbf{x})$, of two different channels to a natural image are shown in Fig. 4.2.2. This figure also shows the responses to a synthetic image with Gaussian statistics. The synthetic image (D) is composed of Gaussian white noise, which has been linearly filtered such that it has a $1/f^2$ power spectrum like natural images. Also, the first order statistics of the synthetic images are the same as those of natural images, which are Gaussian on a logarithmic scale (see Chapter 3.3; Richards, 1982). The response, $a(\mathbf{x})$, with a vertical orientation ($\varphi = 0°$) to a natural image (A) is shown in (B). This channel responds to the vertical structure of the tree-stem. An other channel with an orientation $\varphi = 60°$, shown in (C), responds to the branches of the tree. These responses show spatial structure along the edges of the input image. The same channels, however, produce an unstructured response, shown in (E) and (F), when they respond to the synthetic image (D). We will use synthetic images as comparison for our results. If there is a difference between natural images and the synthetic images, then it must be a result of statistics that are not described by the power spectrum. We will investigate the spatial structure of the responses, $a(\mathbf{x})$, and also dependencies between channels with different orientation or scale.

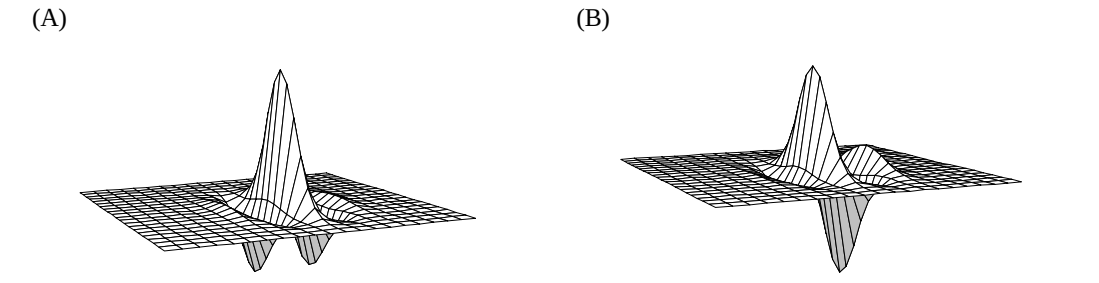

FIGURE 4.2.1: (A) The even receptive field, $e(\mathbf{x})$, and (B) the odd receptive field, $o(\mathbf{x})$, of Eq. 4.2.1. The filters are the real and imaginary part of the inverse Fourier transform of Eq. 4.2.3, with parameters $f_0 = 0.20$ cycles/pixel, $\sigma_1 = 0.54\, f_0$, $\sigma_2 = 0.34\, f_0$, and $\varphi = 0°$, calculated on an image size of 256×256 pixels. Only the central region of 21×21 pixels is shown.

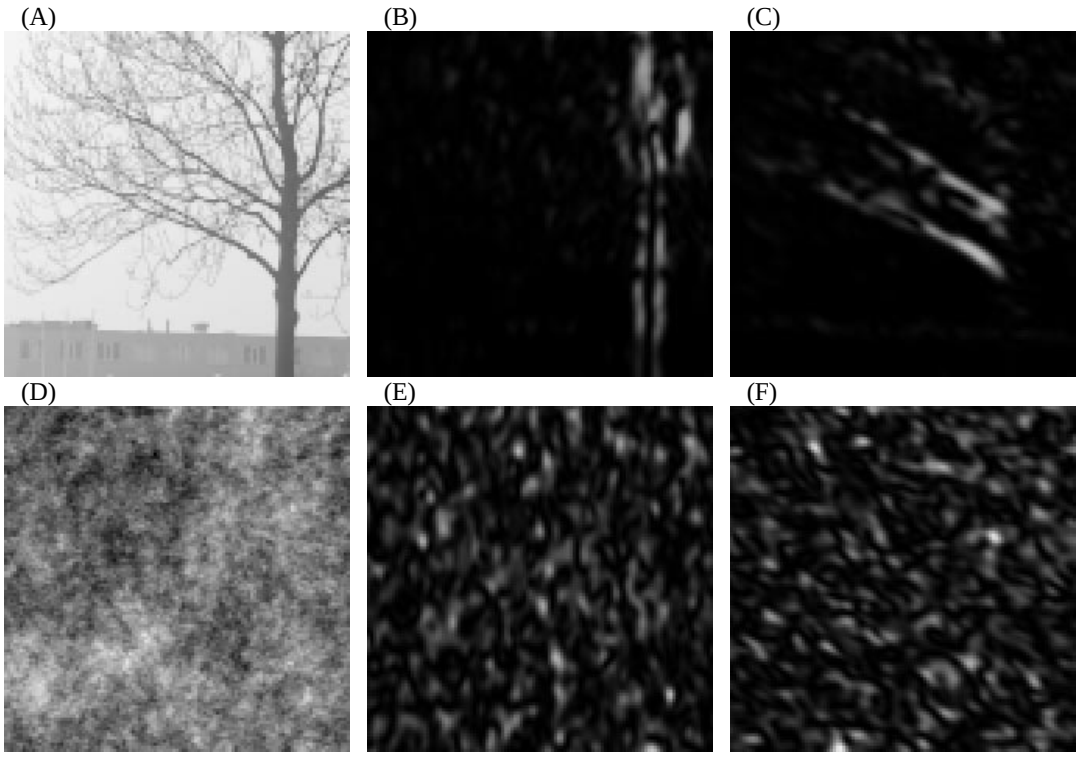

FIGURE 4.2.2: Example of the modelled complex cell responses, $a(\mathbf{x})$. The complex cells are tuned to a frequency $f_0 = 0.2$ cycles/pixel and have a frequency bandwidth of 1.5 octaves and orientation bandwidth of 38°. (A) shows a natural image and (B) the responses of a channel of complex cells with an orientation $\varphi = 0°$. The response of a channel with orientation $\varphi = 60°$ is given in (C). A synthetic image with the same second order statistics as a natural image is given in (D). The response to the synthetic image is given in (E) and (F) for channels with the same parameters as (B) and (C) respectively.

Correlation and mutual information

We want to assess the dependence of complex cells with different parameters. The dependence can be measured by the correlation, which is easy to calculate if we assume translation invariance, such that only the relative distance of the complex cells matters and not their absolute position. We define the correlation of two channels, $a_1(\mathbf{x})$ and $a_2(\mathbf{x})$, over a distance $\boldsymbol{\Delta}$ as:

$$c(\Delta) = \frac{\frac{1}{N}\sum_{\mathbf{x}} (a_1(\mathbf{x}) - \mu_1) \cdot (a_2(\mathbf{x}+\Delta) - \mu_2)}{\sigma_1 \cdot \sigma_2}, \tag{4.2.5}$$

where N is the number of pairs $(a_1(\mathbf{x}), a_2(\mathbf{x}+\boldsymbol{\Delta}))$ in the range $\mathbf{x}$, and

$$\mu_1 = \frac{1}{N}\sum_{\mathbf{x}} a_1(\mathbf{x}), \tag{4.2.6}$$

$$\mu_2 = \frac{1}{N}\sum_{\mathbf{x}} a_2(\mathbf{x}+\Delta), \tag{4.2.7}$$

$$\sigma_1 = \sqrt{\frac{1}{N}\sum_{\mathbf{x}} (a_1(\mathbf{x}) - \mu_1)^2}\ , \tag{4.2.8}$$

$$\sigma_2 = \sqrt{\frac{1}{N}\sum_{\mathbf{x}} (a_2(\mathbf{x}+\Delta) - \mu_2)^2}\ . \tag{4.2.9}$$

The two channels of complex cell responses, $a_1(\mathbf{x})$ and $a_2(\mathbf{x}+\Delta)$, are both calculated from Eq. 4.2.2, but with different parameters. The summation in Eq. 4.2.5 through Eq. 4.2.9 runs over all spatial positions (**x)** of many images, but excludes the borders of the images. More precisely, each pair ($a_1(\mathbf{x})$, $a_2(\mathbf{x}+\Delta)$) of which at least one is located at a position that is 16 pixels or less away from the border of the image is excluded from the summation range. For the present analysis we used a set of 200 well calibrated natural images with a size of 512×512 pixels. The correlation, $c(\Delta)$, of the synthetic Gaussian images was calculated analytically (see Appendix), and the results were checked with numerical calculations on generated synthetic images.

The joint probability, $P(a_1(\mathbf{x}), a_2(\mathbf{x}+\Delta))$, of two complex cell responses can be measured as a two-dimensional histogram. In the present analysis the bins of the histogram are logarithmically spaced, with a bin width of 0.04 $\log_{10}$-units. The histogram is calculated from the same response pairs ($a_1(\mathbf{x})$, $a_2(\mathbf{x}+\Delta)$) that are included in the summation range of the correlation (Eq. 4.2.5). The histogram is normalised, such that

$$\sum_{(a_1,a_2)} P(a_1(\mathbf{x}), a_2(\mathbf{x}+\Delta)) = 1, \tag{4.2.10}$$

where the summation range (a_1, a_2) runs over all histogram bins. The marginal probabilities, $P(a_1(\mathbf{x}))$ and $P(a_2(\mathbf{x}+\Delta))$, are calculated by summation over one of the two coordinates of the histogram $P(a_1(\mathbf{x}), a_2(\mathbf{x}+\Delta))$. Using Shannon's definition of uncertainty (Shannon & Weaver, 1949), the mutual information of the two channels over a distance Δ can be calculated from the normalised histograms as:

$$\begin{aligned} I(\Delta) = &\sum_{(a_1)} P(a_1)\log_2 P(a_1) + \sum_{(a_2)} P(a_2)\log_2 P(a_2) - \\ &- \sum_{(a_1,a_2)} P(a_1,a_2)\log_2 P(a_1,a_2) \end{aligned}\ , \tag{4.2.11}$$

where the suffixes (**x**) and (**x**+Δ) have been dropped for clarity. This measure, $I(\Delta)$**,** shows how much information (measured in bits) one complex cell response, $a_1(\mathbf{x})$, gives about the response, $a_2(\mathbf{x}+\Delta)$, of another complex cell. If the mutual information is zero, the responses of the two cells are independent. The more mutual information, the more dependence between the two. The mutual information of complex cell responses to synthetic images is calculated numerically, using two images of 512×512 pixels that are generated from random Gaussian numbers. Two synthetic images are enough, because their variability is much smaller than that of natural images. This was checked numerically.

RESULTS

The independence between complex cells with different parameters can be analysed very generally from the first order distribution of their responses. This is a result of the fact that the amount of information in a natural image is limited. The representation of the image by the complex cell responses must, therefore, be predictable to a certain degree. When all complex cell responses are independent and not predictable from each other, then they must be predictable by themselves. This means that each complex cells encodes a sparse, but independent image feature. The first order distribution of the modelled complex cell responses is given in Fig. 4.2.3 on a logarithmic scale for natural images and for synthetic images.

The distribution of the responses to natural images is much broader than for synthetic images. This means that complex cells respond to natural images sometimes very strongly, but, most of the time, the response is much smaller than the maximum response. The distribution of the responses to synthetic images is more dense, and, therefore, more often near the maximum response. This means that complex cell responses are much sparser for natural images than for synthetic images. This is not an effect of intensity or contrast differences between the natural and synthetic images, since these parameters affect only the mean, and not the relative width of the first order distribution. That orientation selective representations produce a sparse code for natural images has been reported before (e.g., Daugman, 1989; Field, 1994; Olshausen & Field, 1996). The sparse code is probably a result of the edges that are present in natural images and not in the synthetic images. The responses to edges are sparse, but are they really independent?

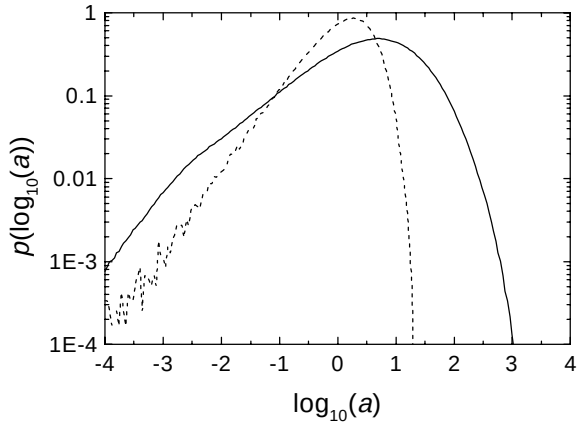

FIGURE 4.2.3: First order probability density distribution of complex cell responses (a, from Eq. 4.2.2) on a logarithmic scale for natural images (solid line) and for synthetic images (dashed line). The parameters of the receptive fields are the same as for Fig 4.2.1. The distribution has a standard deviation of 0.94 $\log_{10}$-units for natural image and 0.56 $\log_{10}$-units for synthetic images.

Spatial dependence

The example of Fig. 4.2.2 shows that, for natural images, there is spatial structure in the response of a single orientational channel of complex cells. The responses appear to be correlated along the edges in the image. This structure can be quantified by the autocorrelation of the channel response. The autocorrelation is computed from Eq. 4.2.5 with $a_1(\mathbf{x}) = a_2(\mathbf{x})$, and is given for natural and synthetic images in Fig. 4.2.4.

We expect that edges in natural images produce high correlation in the direction along the edge. Fig. 4.2.4 shows that, indeed, the autocorrelation of the response to natural images is non-zero over a long distance along the preferred orientation (φ), while it is approximately zero for synthetic images at distances larger than the effective size of the receptive field. Unexpectedly, the autocorrelation of the response to natural images is also non-zero over a long distance in other directions than the preferred orientation (see the elevation of Fig. 4.2.4A compared to Fig. 4.2.4B). Cross sections through Fig. 4.2.4 are shown in Fig. 4.2.5.

The autocorrelation function for natural images in the perpendicular direction, shown in Fig. 4.2.5B, decreases quickly and reaches a plateau value of approximately 0.2 at distances larger than approximately 10 pixels. The autocorrelation function in the preferred direction, shown in Fig. 4.2.5A, decreases only slowly, and reaches the value of 0.2 at a distance of 55 pixels. The autocorrelation function for synthetic images (dashed line in Fig. 4.2.5) does not show the slow decrease in the preferred direction or the plateau in the perpendicular direction. While the first effect can by directly attributed to the presence of edges in natural images, the plateau in the perpendicular direction is more difficult to explain. The increased correlation in the perpendicular direction may be the result of parallel edges or of local contrast variations, which will be discussed later.

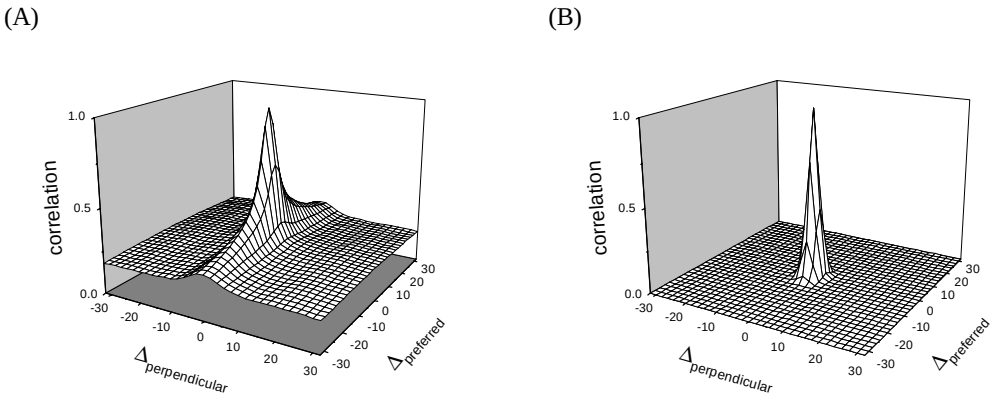

FIGURE 4.2.4: Autocorrelation c(Δ) of complex cell responses to (A) natural images and (B) synthetic images, using the same parameters as in Fig. 4.2.1. The two-dimensional spatial parameter Δ is given in pixels by the coordinate $\Delta_{preferred}$ along the preferred orientation (φ) and the coordinate $\Delta_{perpendicular}$ in the perpendicular direction.

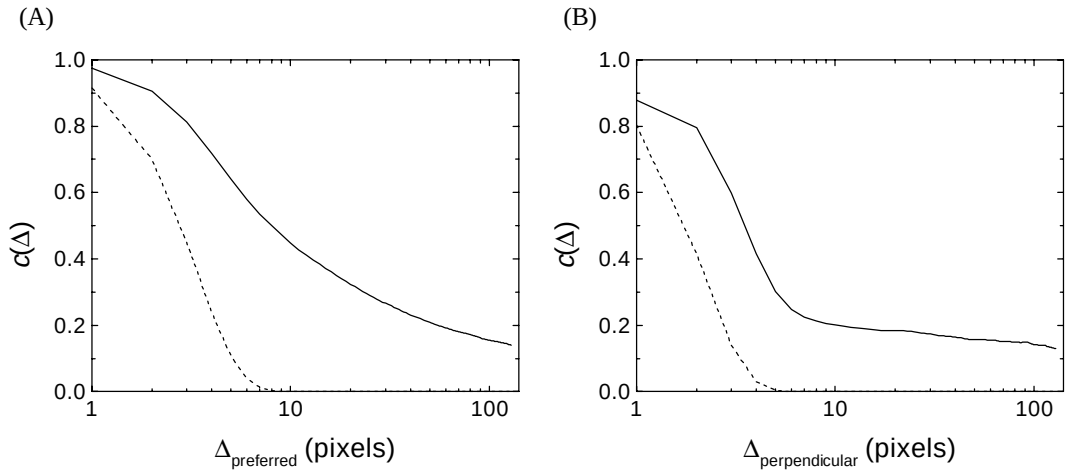

FIGURE 4.2.5: Cross sections of the autocorrelation function of Figure 4.2.4 (A) in the preferred direction and (B) in the perpendicular direction. The solid lines show the autocorrelation for natural images and the dashed line for synthetic images.

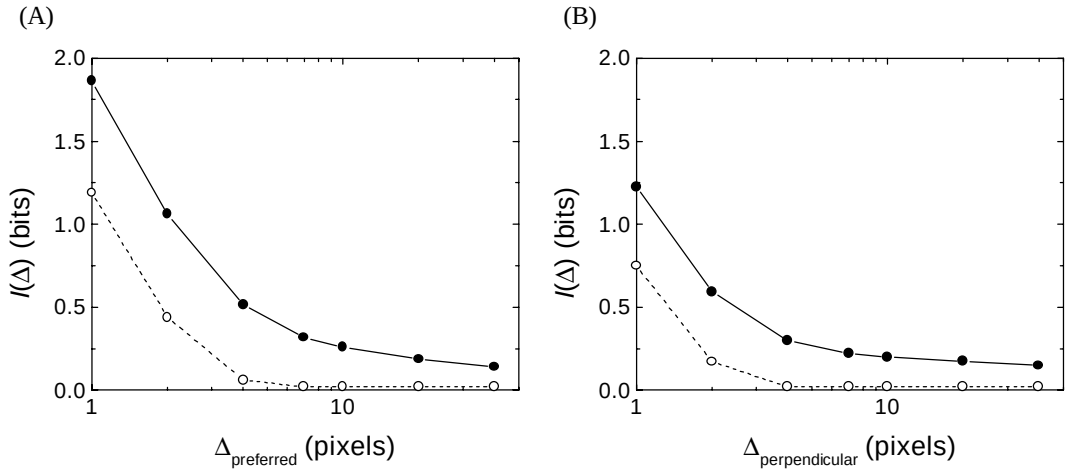

FIGURE 4.2.6: Mutual information of complex cell responses in the same channel that are spaced (A) in the preferred direction and (B) in the perpendicular direction. The solid lines show the results for natural images and the dashed lines for synthetic images.

The mutual information between complex cell responses within the same channel is shown in Fig 4.2.6. At small distances, the mutual information is in the order of 0.5 bit higher for natural images than for synthetic images. The absolute amount of mutual information, however, decreases quickly with distance, and is already less than 0.5 bit at distances larger than the effective size of the receptive field.

The above analysis has been carried out with a fixed parameter setting (see the caption of Fig. 4.2.1). We tested the results with a variety of other settings and found that the results do not critically depend on the parameters, but scale almost linearly with the size and shape of the receptive field. The rescaled results differ less than 0.1 in correlation and 0.1 bit in mutual information after changing the frequency bandwidth or the orientation bandwidth by up to a factor 2, or changing the preferred orientation.

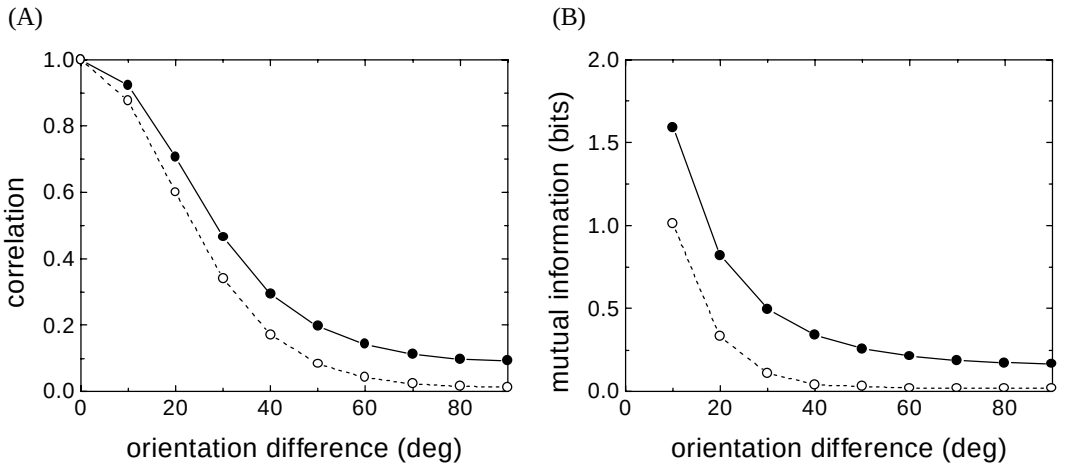

FIGURE 4.2.7: (A) Correlation and (B) mutual information of complex cell responses across channels with different orientation. The solid lines show the results for natural images and the dashed line for synthetic images. The parameters of the channels are the same as in Fig. 4.2.1, but the orientation (φ) of the second channel is varied. Correlation and mutual information are calculated for responses at the same spatial position (i.e., $\mathbf{\Delta} = \mathbf{0}$)

Dependence across orientation

Not only are the responses of complex cells dependent within a channel, but there is also dependence across channels. Fig. 4.2.7A shows the correlation between complex cells that respond to the same spatial location, but with different preferred orientation. The correlation is large for orientation differences smaller than the orientation bandwidth (38°), but becomes closer to 0 for larger orientation differences. This result is approximately the same for natural images and synthetic images. Fig. 4.2.7B shows that the mutual information between complex cell responses is less than 0.4 bits when their preferred orientation differs by more than the orientation bandwidth of 38°. The mutual information between cells with perpendicular preferred orientation is only 0.2 bits. This means that edges with perpendicular orientation appear in natural images quite independently at the same spatial position. Two overlapping perpendicular edges are intrinsically two-dimensional image features, which are rare according to Wegmann and Zetzsche (1990). Our result suggests that such edge crossings are rare just because edges are rare. If perpendicular edges were not independent, but would exclude each other from the same spatial location, then we would have found a negative correlation.

Dependence across scale

A typical property in biological visual systems is that the receptive fields of complex cells are found at many different spatial scales (De Valois, *et al.*, 1982b). However, we may wonder whether such a multiresolution representation is independent. Fig. 4.2.8 shows that the correlation and mutual information between channels of different scale is larger for natural images than for synthetic images. This is probably a result of the edges in natural images, because complex cells of different scale will both respond to the same edge if it falls within their receptive field. The mutual

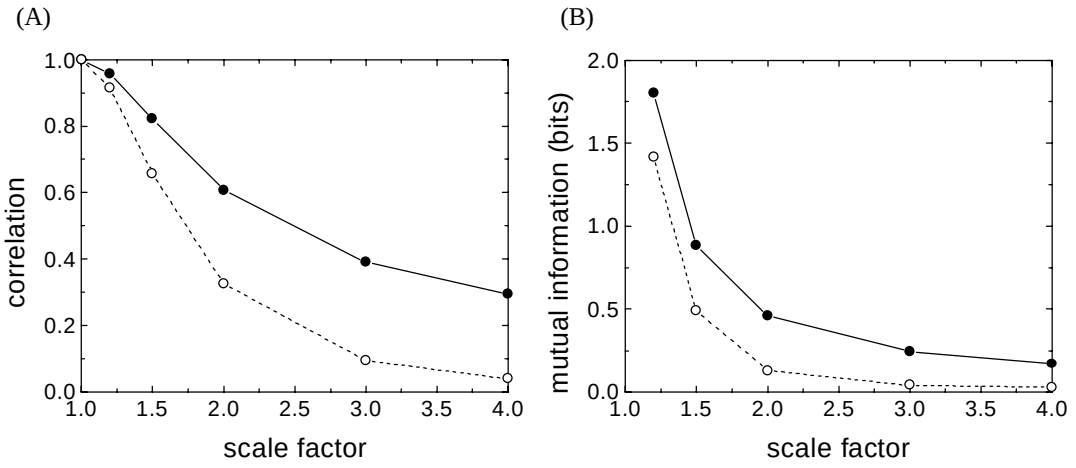

FIGURE 4.2.8: (A) Correlation and (B) mutual information of complex cell responses (at the same spatial position) across channels with different scale. The solid lines show the results for natural images and the dashed line for synthetic images. The parameters of the channels are the same as in Fig. 4.2.1, but the frequency f_0 of the second channel is divided by a scale factor. Note that the scaling affects also σ_1 and σ_2. The analysis is restricted in this calculation to positions that are at least 64 pixels from the image border.

information is, however, less than 0.3 bit for scale factors larger than 2.8, which corresponds to the frequency bandwidth of 1.5 octaves.

LOCAL CONTRAST

The results so far show that the responses of complex cells are more dependent for representations of natural images than for synthetic images. This increase of dependence can be attributed to the presence of edges in natural images. Assuming that independence is a beneficial property for image representation (e.g., Field, 1994), we may ask ourselves the following question: Why does a visual system use an orientation selective representation that explicitly responds to edges? To answer this question we may wonder whether natural images can be represented by independent units that are not orientation selective. As a very general isotropic model we will consider the local variance of the image, which is defined as:

$$a'(\mathbf{x}) = \frac{1}{N}\sum_{\mathbf{x}'}(g(\mathbf{x}') - \mu(\mathbf{x}))^2 , \tag{4.2.12}$$

with

$$\mu(\mathbf{x}) = \frac{1}{N}\sum_{\mathbf{x}'}(g(\mathbf{x}')) , \tag{4.2.13}$$

where the summation range $\mathbf{x}'$ runs over the neighbourhood of 7×7 pixels centred around the spatial position $\mathbf{x}$, and N is the number of pixels in this range ($N = 49$). The effective size (two standard deviations) of this range is 4 pixels, which is approximately the same as the effective size of the complex cell receptive field of Fig. 4.2.1. Here, $g(\mathbf{x}')$ represents the logarithm of the image intensity. Therefore, the local

variance, $a'(\mathbf{x})$, is related to the local contrast of the original image with a linear representation of intensity.

We can now analyse the local variance representation of natural images, $a'(\mathbf{x})$, with the same methods as used for complex cell responses. Local variance, however, is isotropic and does not have an orientation parameter. Fig. 4.2.9A shows the autocorrelation of the local contrast (solid line) together with the autocorrelation of complex cell responses in the preferred direction (dashed line) and in the perpendicular direction (dotted line). The correlation of local contrast is much higher than for the complex cell responses in both directions and extends over a much greater distance. The same is true for the mutual information, which is shown in Fig. 4.2.9B. This means that, for natural images, the units in an isotropic representation are less independent than in an orientation selective representation. For synthetic images, on the contrary, both representations give approximately the same results (not shown), with almost zero correlation and mutual information for distances larger than the effective size of the receptive fields.

The reason for this difference between the representation of natural images and synthetic images can be found from the first order distribution of local variances, which is shown in Fig. 4.2.10. The distribution of local variances is very wide for natural images, whereas it is very narrow for synthetic images, which means that the local variance varies much more for natural images than for synthetic images. This is probably a result of the localised contrasts of the edges that are present in natural images and not in synthetic images. However, not all the variability of contrast is necessarily caused locally by the presence of edges. It may also be caused by large image regions with different contrast. Compare, for example, a textured region with a uniform sky. This large scale variability of contrast is the probable

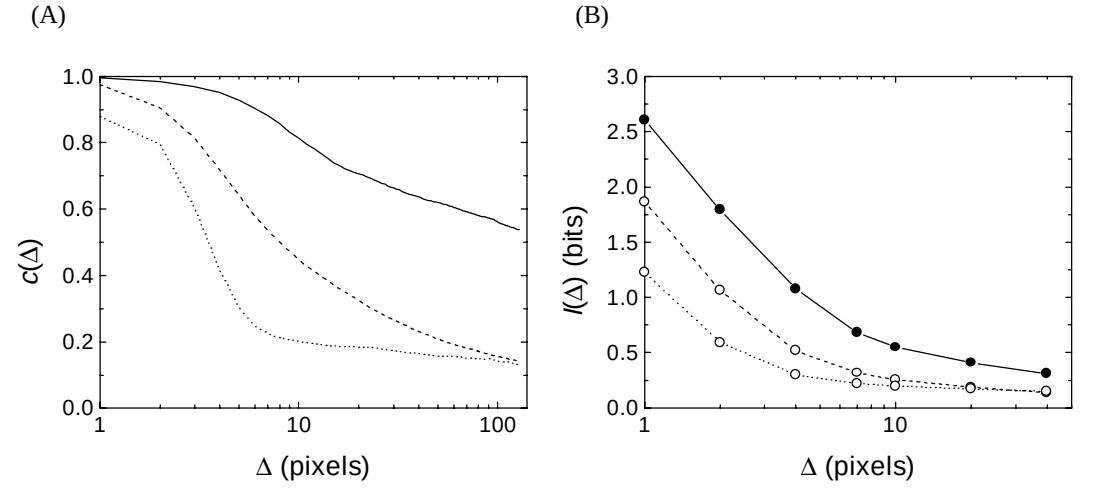

FIGURE 4.2.9: (A) Autocorrelation function of the local variance of natural images and (B) the mutual information between local variances over a distance Δ (solid lines). The local variance is calculated over a 7x7 pixels range. The results from Figs. 4.2.5 and 4.2.6. for complex cell responses to natural images are given for comparison. The results for the preferred orientation is given by the dashed lines and the dotted lines give the results for the perpendicular orientation

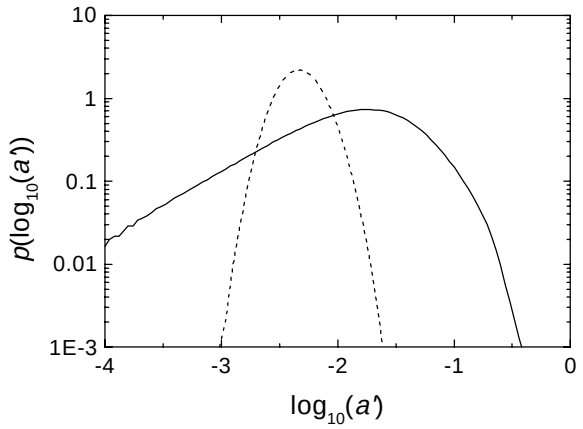

FIGURE 4.2.10: First order probability density distribution of local variance on a logarithmic scale for natural images (solid line) and for synthetic images (dashed line). The distribution has a standard deviation of 0.77 $\log_{10}$-units for natural images and 0.18 $\log_{10}$-units for synthetic images.

cause of the non-zero correlation in the perpendicular direction of the complex cell responses that is shown in Fig. 4.2.5B.

CONCLUSION

We showed that the presence of edges in natural images results in variation of local contrast. This variation of local contrast causes redundancies over large distances in the representation of natural images with isotropic units. The representation with orientation selectivity units, however, is much more independent. Orientation selective complex cell responses have a mutual information that is less than 0.5 bit if they are spaced at least at a distance the size of the effective receptive field. Furthermore, the mutual information between channels of complex cell responses is less than 0.4 bit if they differ at least one bandwidth (38°) in preferred orientation and it is less than 0.3 bit if they are tuned to a scale that is at least one bandwidth (1.5 octaves) apart. This means that complex cell responses are almost independent if their receptive fields do not overlap or when their orientation or scale differ more than one bandwidth.

The result that orientation selectivity increases independence compares well with the results of the independent component analysis of Chapter 4.1, which produces orientation selective filters. The independent component filters are uniformly distributed over space and preferred orientation, which agrees with the present analysis, because we conclude that orientation selective responses are independent if the receptive fields are spaced enough apart or differ enough in preferred orientation. The scaling behaviour of the independent component filters, however, does not compare with the present analysis. The independent component filters are all tuned to the same scale, while the present analysis suggests that filters of different scale can be independent. A possible explanation is that a filter with a large receptive field overlaps not with a single, but with many filters with small receptive fields. The output of each small filter is almost independent, but all the small filters together may well predict the output of the large filter, which is, then, not independent. Compare,

for example, a number of random, independent binary inputs to an exclusive OR operation. The output of the operation is completely defined by all inputs together, but cannot be predicted from each input separately. This example shows that there may still be large redundancies in a system of variables of which each pair is mutually independent. To investigate the consequences for the present analysis, we can extend our methods to systems of more that two complex cell responses. This would, however, require a data set and computing power that grows exponentially with the number of units in the system.

APPENDIX

Here we derive an explicit expression for the correlation ($c(\mathbf{\Delta})$ in Eq. 4.2.5) of complex cell responses to synthetic images. A synthetic image, $g(\mathbf{x})$, of $N \times N$ pixels is composed of filtered Gaussian white noise and is defined by its Fourier transform, $G(\mathbf{f})$, as

$$G(\mathbf{f}) = \frac{1}{|\mathbf{f}|} \cdot \left(\varepsilon(\mathbf{f}) + \varepsilon^*(-\mathbf{f})\right), \tag{4.2.14}$$

where $\varepsilon(\mathbf{f})$ is complex (random phase) zero mean Gaussian white noise. Both coordinates of the two-dimensional spatial frequency vector $\mathbf{f}$ run from $-N/2+1$ to $N/2-1$ with integer steps. This frequency range does not contain the $N/2$ boundary components, which do not contribute significantly to the complex cell responses, but would only complicate the derivation. The combination of $\varepsilon(\mathbf{f})$ and $\varepsilon^*(-\mathbf{f})$ assures that the inverse Fourier transform of $G(\mathbf{f})$ is real. The factor $1/|\mathbf{f}|$ filters the random noise such that the synthetic image has a $1/f^2$ power spectrum, like natural images.

A modelled complex cell response (a) is given by Eq. 4.2.2 as the square modulus of a filtered image, where the filter is complex analytic and given by its frequency response $H(\mathbf{f})$ of Eq. 4.2.3. Applying the inverse Fourier transform, this is written as

$$a = \left| \frac{1}{N^2} \sum_{\mathbf{f}} H(\mathbf{f}) \cdot G(\mathbf{f}) \cdot e^{2\pi i \mathbf{f}^T \mathbf{x}/N} \right|^2, \tag{4.2.15}$$

which can be rewritten with substitution of Eq. 4.2.3 and Eq. 4.2.14 as

$$a = \left| \sum_{\mathbf{f}} \left(\alpha(\mathbf{f}) \cdot \varepsilon(\mathbf{f}) + \alpha(-\mathbf{f}) \cdot \varepsilon^*(\mathbf{f})\right) \right|^2, \tag{4.2.16}$$

with

$$\alpha(\mathbf{f}) = \frac{1}{N^2} \cdot e^{\frac{-(f_1 - f_0)^2}{2\sigma_1^2} + \frac{-f_2^2}{2\sigma_2^2}} \cdot e^{2\pi i \mathbf{f}^T \mathbf{x}/N}, \tag{4.2.17}$$

which depends on six free parameters: the preferred frequency (f_0), the preferred orientation (φ), the bandwidths σ_1 and σ_2, and the two-dimensional spatial position ($\mathbf{x}$).

In the following we use the expectation operator (E) and the properties of the Gaussian statistics of $\varepsilon(\mathbf{f})$, which are in particular:

$$\mathrm{E}[\varepsilon(\mathbf{f})] = 0\,, \tag{4.2.18}$$

because $\varepsilon(\mathbf{f})$ is zero mean,

$$\mathrm{E}[\varepsilon(\mathbf{f}) \cdot \varepsilon(\mathbf{f})] = 0\,, \tag{4.2.19}$$

because $\varepsilon(\mathbf{f})$ is complex with random phase.

$$\mathrm{E}[\varepsilon(\mathbf{f}) \cdot \varepsilon^*(\mathbf{f})] = \mathrm{E}[|\varepsilon(\mathbf{f})|^2] = \sigma_\varepsilon{}^2 \tag{4.2.20}$$

is the variance of $\varepsilon(\mathbf{f})$, and

$$\mathrm{E}[|\varepsilon(\mathbf{f})|^4] = 3\sigma_\varepsilon{}^4 \tag{4.2.21}$$

is a property of the Gaussian distribution. Finally, we use the independence of $\varepsilon(\mathbf{f})$ and $\varepsilon(\mathbf{f}')$ if $\mathbf{f} \neq \mathbf{f}'$, such that

$$\mathrm{E}[f(\varepsilon(\mathbf{f})) \cdot f'(\varepsilon(\mathbf{f}'))] = \mathrm{E}[f(\varepsilon(\mathbf{f}))] \cdot \mathrm{E}[f'(\varepsilon(\mathbf{f}'))]\,, \tag{4.2.22}$$

where $f(\cdot)$ and $f'(\cdot)$ are arbitrary functions. Starting with the expansion of Eq. 4.2.17 and using the above listed properties to eliminate the terms that have zero expectation value, we find after some calculation (involving a few substitutions of the summation range indices) that the expectation value of a is given by

$$\mathrm{E}\,a = 2\sigma_\varepsilon{}^2 \cdot \sum_{\mathbf{f}} |\alpha(\mathbf{f})|^2 \tag{4.2.23}$$

For the expectation value of a^2 we find in the same way that

$$\begin{aligned} \mathrm{E}\,a^2 = \sigma_\varepsilon{}^4 \cdot \Bigg\{ & 2 \cdot \sum_{\mathbf{f}} |\alpha(\mathbf{f})|^4 + 4 \cdot \sum_{\mathbf{f}} |\alpha(\mathbf{f})|^2 \cdot |\alpha(-\mathbf{f})|^2 + \\ & + 8 \cdot \left(\sum_{\mathbf{f}} |\alpha(\mathbf{f})|^2 \right)^2 + 4 \cdot \left| \sum_{\mathbf{f}} \alpha(\mathbf{f}) \cdot \alpha(-\mathbf{f}) \right|^2 \Bigg\} \end{aligned} \tag{4.2.24}$$

With these results we have found the variance ($\sigma_a{}^2$) of a, which is given by

$$\begin{aligned} \sigma_a{}^2 &= \mathrm{E}\,a^2 - (\mathrm{E}\,a)^2 \\ &= \sigma_\varepsilon{}^4 \cdot \Bigg\{ 2 \cdot \sum_{\mathbf{f}} |\alpha(\mathbf{f})|^4 + 4 \cdot \sum_{\mathbf{f}} |\alpha(\mathbf{f})|^2 \cdot |\alpha(-\mathbf{f})|^2 + \\ &\quad + 4 \cdot \left(\sum_{\mathbf{f}} |\alpha(\mathbf{f})|^2 \right)^2 + 4 \cdot \left| \sum_{\mathbf{f}} \alpha(\mathbf{f}) \cdot \alpha(-\mathbf{f}) \right|^2 \Bigg\} \end{aligned} \tag{4.2.25}$$

Consider two sensors of which the responses (a_1 and a_2) are both calculated from Eq. 4.2.16 for the same image, but with different parameters. The parameters appear only in the function $\alpha(\mathbf{f})$ of Eq. 4.2.17, which is distinguished for both sensors as

$\alpha_1(\mathbf{f})$ and $\alpha_2(\mathbf{f})$. The expectation value of the cross product of the two responses (a_1 and a_2) is found from tedious calculation in the same way as the previous calculations.

$$\begin{aligned}\mathbf{E}[a_1 \cdot a_2] = \\ &= \sigma_\varepsilon{}^4 \cdot \left\{ 2\sum_{\mathbf{f}} \left|\alpha_1(\mathbf{f})\right|^2 \cdot \left|\alpha_2(\mathbf{f})\right|^2 + 2\sum_{\mathbf{f}} \left|\alpha_1(\mathbf{f})\right|^2 \cdot \left|\alpha_2(-\mathbf{f})\right|^2 + \right. \\ &+ 2\sum_{\mathbf{f}} \alpha_1(\mathbf{f}) \cdot \alpha_2(-\mathbf{f}) \cdot \alpha_1{}^*(-\mathbf{f}) \cdot \alpha_2{}^*(\mathbf{f}) + \\ &+ 4\sum_{\mathbf{f}} \left|\alpha_1(\mathbf{f})\right|^2 \sum_{\mathbf{f}} \left|\alpha_2(\mathbf{f})\right|^2 + \\ &\left. + 4\left|\sum_{\mathbf{f}} \alpha_1(\mathbf{f}) \cdot \alpha_2{}^*(\mathbf{f})\right|^2 + 4\left|\sum_{\mathbf{f}} \alpha_1(\mathbf{f}) \cdot \alpha_2(-\mathbf{f})\right|^2 \right\}\end{aligned} \tag{4.2.26}$$

The covariance (C) between a_1 and a_2 is herewith given by

$$\begin{aligned}C(a_1, a_2) &= \mathrm{E}[a_1 \cdot a_2] - \mathrm{E}\, a_1 \cdot \mathrm{E}\, a_2 = \\ &= \sigma_\varepsilon{}^4 \cdot \left\{ 2\sum_{\mathbf{f}} \left|\alpha_1(\mathbf{f})\right|^2 \cdot \left|\alpha_2(\mathbf{f})\right|^2 + 2\sum_{\mathbf{f}} \left|\alpha_1(\mathbf{f})\right|^2 \cdot \left|\alpha_2(-\mathbf{f})\right|^2 + \right. \\ &+ 2\sum_{\mathbf{f}} \alpha_1(\mathbf{f}) \cdot \alpha_2(-\mathbf{f}) \cdot \alpha_1{}^*(-\mathbf{f}) \cdot \alpha_2{}^*(\mathbf{f}) + \\ &\left. + 4\left|\sum_{\mathbf{f}} \alpha_1(\mathbf{f}) \cdot \alpha_2{}^*(\mathbf{f})\right|^2 + 4\left|\sum_{\mathbf{f}} \alpha_1(\mathbf{f}) \cdot \alpha_2(-\mathbf{f})\right|^2 \right\}\end{aligned} \quad , \tag{4.2.27}$$

with

$$\begin{aligned}&\sum_{\mathbf{f}} \alpha_1(\mathbf{f}) \cdot \alpha_2(-\mathbf{f}) \cdot \alpha_1{}^*(-\mathbf{f}) \cdot \alpha_2{}^*(\mathbf{f}) = \\ &= \sum_{\mathbf{f}} \mathrm{Re}\{\alpha_1(\mathbf{f}) \cdot \alpha_2(-\mathbf{f}) \cdot \alpha_1{}^*(-\mathbf{f}) \cdot \alpha_2{}^*(\mathbf{f})\}\end{aligned} \tag{4.2.28}$$

more convenient to calculate. Finally, the correlation between a_1 and a_2 is given by

$$c(a_1, a_2) = \frac{C(a_1, a_2)}{\sigma_{a_1} \cdot \sigma_{a_2}}, \tag{4.2.29}$$

with $\sigma_{a_1}{}^2$ the variance ($\sigma_a{}^2$ in Eq. 2.4.25) of a_1 and $\sigma_{a_2}{}^2$ the variance of a_2.